# Filtres et contrôles

## Filtres en entrée sur les données aberrantes ou douteuses

### Objectif

Un premier niveau de filtre s’assurera de vérifier la pertinence des données insérées en entrée. Ces données pourront être qualifiées « de météorologiquement aberrantes ou douteuses ».

### Champ d’action

Ceux-ci seront appliqués à **chaque nouvelle insertion de données en base**, qu’elle soit insérée manuellement ou automatiquement. Par ailleurs, elle concernera **autant de niveau d’agrégations qu’il y est possible d’insérer des données**. Il conviendra donc d’adapter au mieux les seuils pour chaque niveau d’agrégation et cela principalement pour les données de type ‘cumul ‘.

### Filtres sur un paramètre

De manière générale, sauf pour les paramètres cumulés, on pourra prendre les mêmes seuils d’un niveau d’agrégation à un autre. Si un filtre est manquant pour un paramètre :

* A défaut, on peut prendre les valeurs limites du capteur.
* Ou, tout simplement en discuter entre membre du groupe de travail afin de trouver des valeurs aberrantes.

**Nomenclature :**

FSXY\_Z :

* X étant le type de contrôle (A : aberrant, D : douteux)
* Y étant le niveau d’agrégation (E : élémentaire, H : horaire, Q : quotidien, M : mensuel, A : annuel)
* Z étant l’abréviation du paramètre contrôlé

**Pour l’ensemble des tables :**

|  |  |  |  |
| --- | --- | --- | --- |
| **Paramètre** | **Nom du paramètre dans la table** | **Aberrant** | **Douteux** |
| **FS{X}{Y}\_T** : Température |  | Entre -20/45°C | Pour la température : Entre -10 et 35°C |
| **FS{X}{Y}\_TD** : Température point de rosée |  |
| **FS{X}{Y}\_TM10 :** Température à 10 cm dans le sol |  | Entre – 8°C et +42°C | - |
| **FS{X}{Y}\_TM20 :** Température à 20 cm dans le sol |  | Entre -6°C et +38°C | - |
| **FS{X}{Y}\_TM50 :** Température à 50 cm dans le sol |  | Entre -5°C et +40°C | - |
| **FS{X}{Y}\_UVINDICE :** Indice UV |  | Entre 0 et 20 |  |
| **FS{X}{Y}\_DD :** Direction du vent vent moyen |  | Entre 0 et 360° |  |
| **FS{X}{Y}\_DXI :** Direction du vent de la rafale max |  |  |
| **FS{X}{Y}\_FF :** Force du vent moyen |  | Entre 0 et 70 m/s |  |
| **FS{X}{Y}\_FXI :** Force des rafales |  | Entre 0 et 100 m/s |  |
| **FS{X}{Y}\_U :** Humidité |  | Entre 0 et 110% | Entre 5 et 100% |
| **FS{X}{Y}\_UN :** Humidité minimale |  |
| **FS{X}{Y}\_UX :** Humidité maximale |  |
| **FS{X}{Y}\_PMER :** Pression au niveau de la mer |  | Entre 850 et 1060 hPa | - |
| **FS{X}{Y}\_PSTAT :** Pression station |  | Entre 600 et 1060 hPa | - |
| **FS{X}{Y}\_H{X}** : Heure des extrêmes, avec X le paramètre |  | Doit être compris dans [H-60mn,H] |  |

**Table aggrégée :**

|  |  |  |  |
| --- | --- | --- | --- |
| **Paramètre** | **Nom du paramètre dans la table** | **Aberrant** | **Douteux** |
| **FS{X}{Y}\_RRX :** Cumul de précipitations sur Xh |  |  | 5min : Entre 0 et 50mm  1h Entre 0 et 250mm  24h : Entre 0 et 1500mm  1 mois : Entre 0 et 5000mm |
| **FS{X}{Y}\_INS :** Durée horaire de l’insolation |  | 1h : Entre 0 et 60min  24h : Entre 0 et 960min (<18h)  1 mois : Entre 0 et 30 000min |  |
| **FS{X}{Y}\_GLO :** Rayonnement horaire global |  | 1h : Entre 0 et 500 J/cm2  24h Entre 0 et 5000 J/cm2 |  |
| **FS{X}{Y}\_INFRAR :** Rayonnement infra-rouge horaire |  | 1h : Entre 0 et 300 J/cm2 |  |

### Filtre inter-paramètre

Nomenclature :

* **FCAH\_X : Filtre Contrainte Absolue Horaire**
* **FCAQ\_X : Filtre Contrainte Absolue Quotidienne**

**Table agg\_hour :**

|  |  |  |
| --- | --- | --- |
| **Paramètre** | **Nom du paramètre dans la table** | **Filtre** |
| **FCAHT\_TN** | **T(H) >= TN(H)** | Température supérieure à la température minimale |
| **FCAHT\_TX** | **T(H) <= TX(H)** | Température inférieure à la température maximale |
| **FCAHU\_UN** | **U(H) >= UN(H)** | Humidité supérieure à l’humidité minimale |
| **FCAHU\_UX** | **U(H) <= UX(H)** | Humidité inférieure à l’humidité maximale |
| **FCAHPMER\_PSTAT** | **PMER(H) >= PSTAT(H)** | Pression mer supérieure à la pression station |

**Table agg\_day :**

|  |  |  |
| --- | --- | --- |
| **Paramètre** | **Nom du paramètre dans la table** | **Filtre** |
| **FCAQTX\_TN** | **TX(Q) >= TN(Q)** | Température maximale supérieure à la température minimale |
| **FCAQUX\_UN** | **UX(Q) >= UN(Q)** | Humidité maximale supérieure à l’humidité minimale |

### Action sur les données filtrées

Si la donnée est qualifiée :

* **Aberrante :** Ces données seront **automatiquement supprimées** de la base opérationnelle **et transférée vers une base annexe**. Les tables de niveau d’agrégation supérieur se verront impactées par ce manque de données (ex : cumuls, min/max..).
* **Douteuse :** Ces données seront **flaggées** dans la base opérationnelle **et dupliquée vers une base annexe**. Les tables de niveau d’agrégation supérieure se verront impactées par ce manque de données. Ils nécessiteront une validation à postériori.

## Trigger simple

Il s’agit d’un contrôle d’intégrité et de mise en cohérence automatique entre plusieurs paramètres.

### Temporalité du contrôle

Ces contrôles devront être effectués à chaque modification (au sens large) des données

### Champ d’action du contrôle

Ce contrôle se fera sur **l’ensemble des tables de la BDD**. En effet, toute table de la BDD peut être modifiée et la modification de la valeur d’un paramètre en « valeur manquante » doit pouvoir être répercutée sur les paramètres connexes. Par ailleurs, la modification d’une valeur doit pouvoir modifier la valeur des paramètres calculés/agrégés (ex : modification de rain\_sum1h entrainera la modification de rain\_sum3h).

### Triggers de mise à manquant

|  |  |
| --- | --- |
| **Si le paramètre est manquant** | **Les paramètres suivants sont mis à manquant :** |
| Vitesse du vent | Direction du vent + heures associées |
| Rafales | Direction de la rafale + heures associées |
| Température minimale | Heure de la température minimale |
| Température maximale | Heure de la température maximale |
| Température | Température minimale, Heure de la TN, Température maximale, Heure de la TX |
| Humidité minimale | Heure de l’humidité minimale |
| Humidité maximale | Heure de l’humidité maximale |
| Humidité | Humidité minimale, Heure de l’humidité minimale, Humidité maximale, Heure de l’humidité maximale |
| Pression au niveau de la mer | Pression au niveau de la mer minimale et heures associées |

## Contrôles à postériori

## Temporalité du contrôle et bases concernées

Ces contrôles peuvent être lancés :

* Quotidiennement sur les données temps réel inférieures à 60j pour les tables ‘données élémentaires’, H et Q.
* À la demande sur des données plus anciennes (ex : intégration d’archives) sur une ou plusieurs tables ‘données élémentaires’, H et Q.

### Contrôles dans agg\_hour

**Contrôle simple :**

**Nomenclature :**

CS{X}\_{Y} :

* X étant le niveau d’agrégation
* Y le paramètre

|  |  |  |  |
| --- | --- | --- | --- |
| **Code** | **Paramètre** | **Nom du paramètre en base** | **Contrôle associé (Aberrant)** |
| CSH\_INS | Durée d’insolation |  | < à 3mn (offset éventuel du capteur) entre 21UTC et 3UTC |
| CSH\_GLO | Rayonnement global horaire |  | < = à 10 J/cm2 (offset éventuel du capteur) entre 21UTC et 3UTC |
| CSH\_HEXT | Heures des extrêmes |  | Les heures des extrêmes horaires doivent appartenir à l’intervalle [h-59mn, h] |
| CSH\_UV | Rayonnement UV |  | <= 0.02J/cm2 entre 21UTC et 3UTC |

**Principe du contrôle temporel :**

* P(h) la valeur du paramètre p à l’observation h
* P(h-1) et p(h+1) les valeurs de ce paramètre aux observations des heures précédentes et suivantes. On calcule une valeur estimée du paramètre p\*(h) = ((p(h-1) +p(h+1))/2

L’écart en valeur absolue entre valeur estimée et valeur observée | p\*(h) – p(h) | doit être inférieur à un seuil fixe pour chaque paramètre.

**Nomenclature :**

CT{X}\_{Y}

|  |  |  |  |
| --- | --- | --- | --- |
| **Code** | **Paramètre** | **Nom du paramètre en base** | **Contrôle associé (douteux)** |
| CTH\_T | Température |  | Seuil fixe : écart < 10°C |
| CTH\_TX ou CTH\_TN | Températures extrêmes **de l’heure** |  | Seuil fixe : écart < 10°C |
| CTH\_PMER | Pression mer |  | Seuil fixe : écart <5hPa |
| CTH\_PMERMIN | Pression de la mer minimale |  | Seuil fixe : écart <5hPa |
| CTH\_FF | Vent moyen |  | Seuil fixe : écart < 10m/s |
| CTH\_FX | Rafales |  | Seuil fixe : écart < 15m/s |
| CTH\_U | Humidité |  | Seuil fixe : écart < 50% |
| CTH\_UX ou CTH\_UN | Humidités extrêmes de l’heure |  | Seuil fixe : écart < 50% |

Voir annexe B et C pour plus de contrôles temporels.

**Contrôles de capteurs bloqués :**

* On ne peut pas avoir plus de N valeurs consécutives (horaires) égales.
* N est dépendant du paramètre

**Nomenclature :**

CCB\_{Y} avec Y le capteur.

|  |  |  |  |
| --- | --- | --- | --- |
| **Code** | **Capteur bloqué** | **Paramètre contrôlé** | **N** |
| CCB\_ANEMO | Anémomètre bloqué | Rafales  Vent moyen | N = 12  N = 24 |
| CCB\_GIR | Girouette bloquée | Direction du vent | N=12 |
| CCB\_THERM | Thermomètre bloqué | Température | N=12 |

**Contrôle de cohérence inter-paramètres**

**Nomenclature :**

CCIP{X}\_{Y}\_{Z} : Y et Z étant les 2 paramètres sur lesquels la cohérence est testée

|  |  |  |
| --- | --- | --- |
| **Code** | **Paramètres contrôlés** | **Contrôles (aberrant)** |
| CCIPH\_TN\_TX | Températures extrêmes | Erreur si TX(H) – TN(H) > 10° |
| CCIPH\_FF\_DD | Vent / Rafales et direction | DD = 0 => FF=0 |
| CCIPH\_FF\_FX | Vent et rafales | FF(H) – 1.1m/s <= Rafales(H) |
| CCIPH\_PMERMIN\_PMER | Pression mer | Erreur si PMERMIN > PMER |
| CCIPH\_T\_TD | Température et température de rosée | Erreur si T > Td + 0.1 |

## Table annexe

Une base annexe se verra remonter plusieurs informations sur les bugs rencontrés lors des filtres et contrôles à postériori. Elle permettra :

* De connaître si un capteur est bloqué
* Si des valeurs aberrantes sont détectées
* Si des valeurs douteuses sont détectées. Celui-ci doit permettre de valider ou d’invalider la valeur dans une table concernée.

## Résumé

La BDD sera soumise à plus contrôles :

* Des filtres exécutés à chaque insertion de données en base et à tous les niveaux d’agrégations. Ceux-ci s’assurent de la pertinence de la donnée insérée en la qualifiant de pertinente, douteuse ou aberrante.
* Des triggers permettant d’assurer la cohérence des données en base. Ceux-ci sont exécutés à chaque insertion de données dans la base.
* Des contrôles simples s’effectuant à postériori sur quelques paramètres et sur une certaine plage horaire. Ils ont pour objectif de qualifier la donnée d’aberrante.
* Un contrôle temporel s’effectuant à postériori et ont pour objectif de qualifier la donnée de douteuse.
* Un contrôle de capteurs bloqués s’effectuant à postériori et à titre informatif.
* Un contrôle de cohérence inter-paramètres effectué à postériori ayant pour objectif de qualifier la donnée d’aberrante.

Ainsi, si les filtres ont pour vocation à être effectué après chaque insertion de la donnée en base, les contrôles se feront quant à eux quotidiennement sur les données temps réel inférieures à 60j pour les tables ‘données élémentaires’, H et Q.

Par ailleurs, ces filtres et contrôles ont pour objectif de qualifier une donnée de pertinente, d’aberrante ou de douteuse. L’effet de cette qualification sur les données est la suivante :

* **Aberrante :** Non insertion de la donnée en base opérationnelle mais insertion en base annexe
* **Douteuse :** Insertion de la donnée en base opérationnelle et en base annexe. Cependant, la donnée en base sera flaggée de douteuse et nécessite une validation humaine. Ainsi, celle-ci devra aboutir soit sur une qualification en tant que pertinente, soit aberrante.

Cette information se propagera aux tables de niveaux d’agrégation supérieur (un cumul horaire de pluie douteuse entraînera un cumul annuel de pluie douteux...) et pourra entraîner la modification des valeurs (un cumul horaire manquant entraînera une modification des cumuls supérieurs). Ces répercutions (notamment d’une qualification d’aberrante) devront être clairement définies.

Une table annexe permettra de recenser tous les contrôles et filtres qui auront été déclenchés. Celle-ci permettra de prendre directement connaissance :

* De l’identifiant du filtre/contrôle
* De la valeur (ou des valeurs pour les cohérences inter-paramètres) concerné.e.s.
* Du/des niveaux d’agrégations des valeurs
* De la date (ou des dates) concernées

Chaque ligne de la table annexe pourra contenir au maximum 2 paramètres, dates ou valeurs (cohérence inter-paramètre). L’archivage de toutes ces informations permettra d’alimenter une interface de visualisation afin d’en faciliter la lecture, le suivi et les actions (ex : qualification d’une donnée douteuse en aberrante ou pertinente)

# ANNEXE

## (MF) Contrôle fourchette

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Type de controle** | **Libellé paramètre BUFS** | **Limite infèrieure** | **Limite supérieure** | **Unité** | **consequence** |
| Contrôle fourchette | Pression mer | 850 | 1050 | Hpa | paramètre aberrant pour ce contrôle |
| Contrôle fourchette | Pression station | 850 | 1050 | Hpa | paramètre aberrant pour ce contrôle |
| Contrôle fourchette | Temperature de l'air minute | LiminfT voir feuille annexe | LimsupT  Voir feuille annexe | °C | paramètre aberrant pour ce contrôle |
| Contrôle fourchette | maximum horaire de Tair | LiminfT voir feuille annexe | LimsupT  Voir feuille annexe | °C | paramètre aberrant pour ce contrôle |
| Contrôle fourchette | minimum horaire de Tair | LiminfT voir feuille annexe | LimsupT  Voir feuille annexe | °C | paramètre aberrant pour ce contrôle |
| Contrôle fourchette | Maximum horaire de Tair sur 24h | LiminfT voir feuille annexe | LimsupT  Voir feuille annexe | °C | paramètre aberrant pour ce contrôle |
| Contrôle fourchette | Minimum horaire de Tair sur 24h | LiminfT voir feuille annexe | LimsupT  Voir feuille annexe | °C | paramètre aberrant pour ce contrôle |
| Contrôle fourchette | Humidité relative minute | 0 | 105 | % | paramètre aberrant pour ce contrôle |
| Contrôle fourchette | Humidité relative minute | 5 | 100 | % | paramètre douteux |
| Contrôle fourchette | minimum horaire de U | 0 | 105 | % | paramètre aberrant pour ce contrôle |
| Contrôle fourchette | minimum horaire de U | 5 | 100 | % | paramètre douteux |
| Contrôle fourchette | vent moyen 10 min : direction | 0 | 360 | ° | paramètre aberrant pour ce contrôle |
| Contrôle fourchette | vent moyen 2 min : direction | 0 | 360 | ° | paramètre aberrant pour ce contrôle |
| Contrôle fourchette | vent moyen 10 min : vitesse | 0 | 70 | m/s | paramètre aberrant pour ce contrôle |
| Contrôle fourchette | vent instantané max 1 min : vitesse | 0 | 80 | m/s | paramètre aberrant pour ce contrôle |
| Contrôle fourchette | vent instantané maximum horaire : vitesse | 0 | 80 | m/s | paramètre aberrant pour ce contrôle |
| Contrôle fourchette | maximum horaire du vent moyen 10 min : vitesse | 0 | 70 | m/s | paramètre aberrant pour ce contrôle |
| Contrôle fourchette | Force du vent moyen 3s minimal sur 10 minutes | 0 | 80 | m/s | paramètre aberrant pour ce contrôle |
| Contrôle fourchette | Force de vent moyen 3s maximal sur 10 minutes | 0 | 80 | m/s | paramètre aberrant pour ce contrôle |
| Contrôle fourchette | maximun horaire du vent myen 10 min : vitesse | 0 | 80 | m/s | paramètre aberrant pour ce contrôle |
| Contrôle fourchette | cumul précipitations en 1 min | 0 | 40 | mm | paramètre aberrant pour ce contrôle |
| Contrôle fourchette | cumul précipitations en 6 min | 0 | 50 | mm | paramètre aberrant pour ce contrôle |
| Contrôle fourchette | cumul précipitations en 1H | 0 | 200 | mm | paramètre aberrant pour ce contrôle |
| Contrôle fourchette | cumul précipitations en 3H | 0 | 350 | mm | paramètre aberrant pour ce contrôle |
| Contrôle fourchette | cumul précipitations en 6H | 0 | 350 | mm | paramètre aberrant pour ce contrôle |
| Contrôle fourchette | cumul précipitations en 12H | 0 | 600 | mm | paramètre aberrant pour ce contrôle |
| Contrôle fourchette | cumul précipitations en 24H | 0 | 1200 | mm | paramètre aberrant pour ce contrôle |
| Contrôle fourchette | durée d'insolation dans la minute | 0 | 1 | min | paramètre aberrant pour ce contrôle |
| Contrôle fourchette | durée d'insolation dans les 6 min | 0 | 6 | min | paramètre aberrant pour ce contrôle |
| Contrôle fourchette | durée d'insolation dans l'heure | 0 | 60 | min | paramètre aberrant pour ce contrôle |
| Contrôle fourchette | Rayonnement global dans la minute | -20 | 500 | J/m2 | paramètre aberrant pour ce contrôle |
| Contrôle fourchette | Rayonnement global dans les 6 min | -20 | 500 | J/m2 | paramètre aberrant pour ce contrôle |
| Contrôle fourchette | Rayonnement global dans l'heure | -20 | 500 | J/m2 | paramètre aberrant pour ce contrôle |
| Contrôle fourchette | rayonnement diffus dans la minute | -20 | 500 | J/m2 | paramètre aberrant pour ce contrôle |
| Contrôle fourchette | rayonnement diffus dans les 6 min | -20 | 500 | J/m2 | paramètre aberrant pour ce contrôle |
| Contrôle fourchette | Rayonnement diffus dans l'heure | -20 | 500 | J/m2 | paramètre aberrant pour ce contrôle |
| Contrôle fourchette | Rayonnement direct dans la minute | -20 | 500 | J/m2 | paramètre aberrant pour ce contrôle |
| Contrôle fourchette | Rayonnement direct dans les 6 min | -20 | 500 | J/m2 | paramètre aberrant pour ce contrôle |
| Contrôle fourchette | Rayonnement direct dans l'heure | -20 | 500 | J/m2 | paramètre aberrant pour ce contrôle |

## (MF) Contrôle temporel : gradient min

|  |  |  |  |
| --- | --- | --- | --- |
| **durée entre valeurs successives** | **Test proposé** | **Nombre de valeurs successives N** | **Conséquence** |
| 1 min | |Pt-Pt-1|<0,0001 | 30 | Donnée N douteuse |
| 1 min | |Pt-Pt-1|<0,0001 | 30 | Donnée N douteuse |
| 1 min | |Tt-Tt-1|<0,0001 | 30 | Donnée N douteuse |
| 6 min | |Tt-Tt-1|<0,0001 | 10 | Donnée N douteuse |
| 1H | |Tt-Tt-1|<0,0001 | 3 | Donnée N douteuse |
| 1H | |Tt-Tt-1|<0,0001 | 3 | Donnée N douteuse |
| 1 min | U<81 et |Ut-Ut-1|<0,1 | 30 | Donnée N douteuse |
| 6 min | U<81 et |Ut-Ut-1|<0,1 | 20 | Donnée N douteuse |
| 1H | U<81 et |Ut-Ut-1|<0,1 | 3 | Donnée N douteuse |
| 1H | U<81 et |Ut-Ut-1|<0,1 | 3 | Donnée N douteuse |
| 1 min | ff>0 et |ddt-ddt-1|<0,1 | 30 | Donnée N douteuse |
| 6 min | ff>0 et |ddt-ddt-1|<0,1 | 20 |
| 1H | ff>0 et |ddt-ddt-1|<0,1 | 3 |
| 1 min | ff>0 et |ddt-ddt-1|<0,1 | 30 | Donnée N douteuse |
| 6 min | ff>0 et |ddt-ddt-1|<0,1 | 20 |
| 1H | ff>0 et |ddt-ddt-1|<0,1 | 3 |
| 1 min | ff>0 et |ddt-ddt-1|<0,1 | 30 | Donnée N douteuse |
| 6 min | ff>0 et |ddt-ddt-1|<0,1 | 20 |
| 1H | ff>0 et |ddt-ddt-1|<0,1 | 3 |
| 1 min | ff>0 et |ddt-ddt-1|<0,1 | 30 | Donnée N douteuse |
| 6 min | ff>0 et |ddt-ddt-1|<0,1 | 20 |
| 1H | ff>0 et |ddt-ddt-1|<0,1 | 3 |
| 1 min | ff>0 et |ddt-ddt-1|<0,1 | 30 | Donnée N douteuse |
| 6 min | ff>0 et |ddt-ddt-1|<0,1 | 20 |
| 1H | ff>0 et |ddt-ddt-1|<0,1 | 3 |

## (MF) Contrôle temporel : gradient max

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **paramètre** | **Libellé paramètre BUFS** | **durée entre valeurs succesives** | **Test proposé** | **consequence** |
| Psta | Pression station | 1 min | |Pt-Pt-1|>4 | Donnée douteuse si donnée -1 correcte |
| 6 min | |Pt-Pt-1|>4 |
| 1H | |Pt-Pt-1|>5 |
| Pmer | Pression mer | 1 min | |Pt-Pt-1|>4 | Donnée douteuse si donnée -1 correcte |
| 6 min | |Pt-Pt-1|>4 |
| 1H | |Pt-Pt-1|>5 |
| T | Température de l'air minute | 1 min | |Tt-Tt-1|>3 | Donnée douteuse si donnée -1 correcte |
| 6 min | |Tt-Tt-1|>8 |
| 1H | |Tt-Tt-1|>18 |
| TSYN | Température de l'air minute | 1H | |Tt-Tt-1|>18 | Donnée douteuse si donnée -1 correcte |
| U | Humidité relative minute | 1 min | |Ut-Ut-1|>20 | Donnée douteuse si donnée -1 correcte |
| 6 min | |Ut-Ut-1|> 30 |
| 1H | |Ut-Ut-1|>50 |
| USYN | Humidité relative minute | 1H | |Ut-Ut-1|>50 | Donnée douteuse si donnée -1 correcte |